BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 8: e136697
doi: 10.3897/biss.8.136697

OPEN ACCESS

Conference Abstract

The European MAREGRAPH Project: Enhancing Marine Data Interoperability through Semantic Knowledge Graphs

Giorgia Lodi‡, Alessandro Russo‡, Joanna Goley§, Marc Portier§, Katrina Exter§

‡ Institute of Cognitive Sciences and Technologies - Italian National Research Council (CNR-ISTC), Rome, Italy
§ Flanders Marine Institute (VLIZ), Oostende, Belgium

Corresponding author: Alessandro Russo (alessandro.russo@cnr.it)

Received: 09 Sep 2024 | Published: 10 Sep 2024

Citation: Lodi G, Russo A, Goley J, Portier M, Exter K (2024) The European MAREGRAPH Project: Enhancing Marine Data Interoperability through Semantic Knowledge Graphs. Biodiversity Information Science and Standards 8: e136697. https://doi.org/10.3897/biss.8.136697

Abstract

The availability of accurate and up-to-date information on marine species—ranging from nomenclature and taxonomy to species occurrences—is fundamental for several stakeholders, from marine scientists to policymakers. Today, this information is provided by well-known authoritative systems, including the World Register of Marine Species (WoRMS*[1]), a comprehensive register of names of marine organisms and related information; Marine Regions*[2], a register of georeferenced information on marine place names and areas; and the European Ocean Biogeographic Information System (EurOBIS*[3]), where species occurrences in European waters are recorded. These platforms ensure marine biodiversity data is accessible and interoperable, supporting various applications and use cases for fostering advancements in taxonomy, marine biology, ecology and environmental science.

The MAREGRAPH project*[4] has the ambition to further increase the findability, accessibility, interoperability and reusability of these foundational high-value datasets, by creating and publishing an open knowledge graph (KG) on marine biodiversity data through the semantic uplifting of the involved datasets, relying on the Semantic Web stack and its standards (Fig. 1).

© Lodi G et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

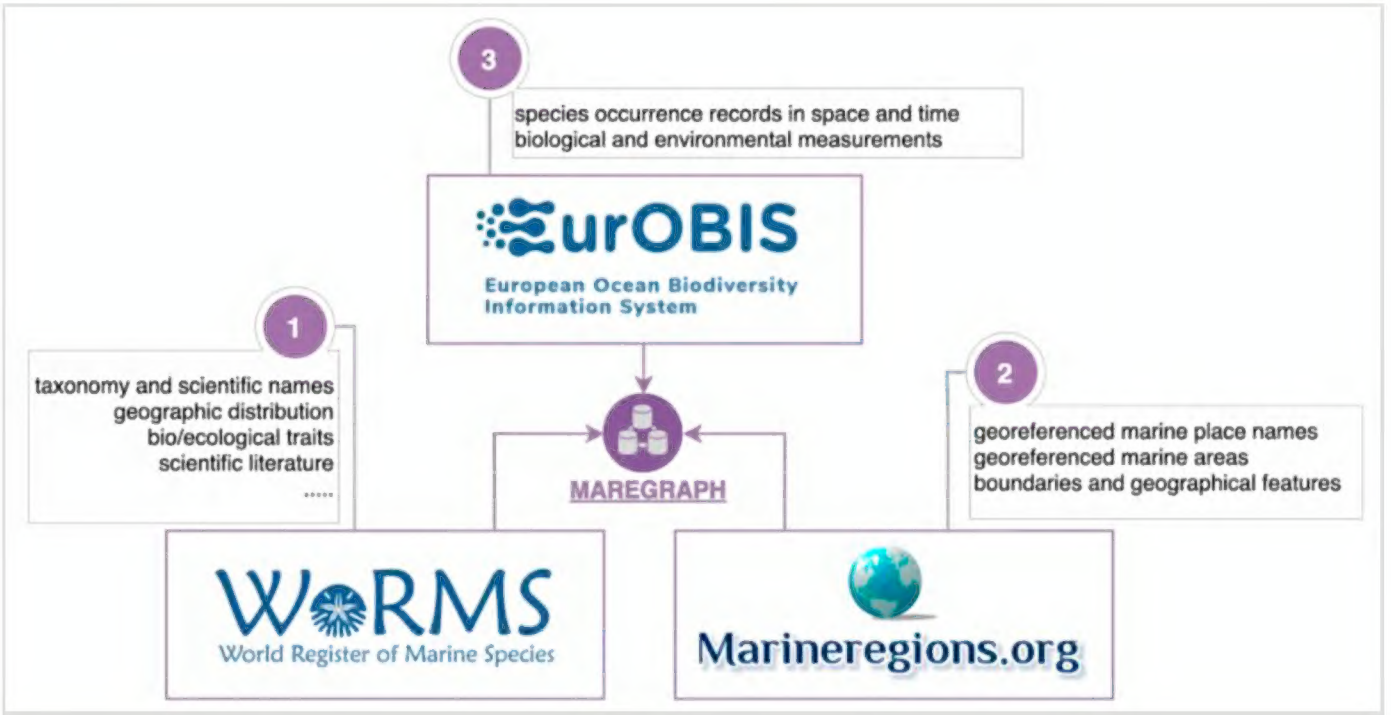

Figure 1.

Data sources and knowledge domains for the MAREGRAPH knowledge graph (CC BY-SA 4.0)

This requires:

1. the definition and publication of a network of formal, reusable and extensible ontologies and controlled vocabularies for representing and linking together marine biodiversity information, as a basis for semantic interoperability;
2. the set up of a cost-effective, flexible and scalable data architecture to publish the data as Linked Open Data and provide access through standard APIs, from a SPARQL endpoint to Linked Data Event Streams (LDES), enabling technical interoperability and serving various use cases and data access needs.

The methodology adopted for defining and publishing the KG combines the consolidated principles of the Open Standards for Linked Organisations (OSLO) framework—which defines the governance structure and a open process for developing semantic data standards—with the well-established agile and collaborative ontology development workflow of the eXtreme Design methodology (Presutti 2009), based on ontology design patterns (ODPs).

Also leveraging the experience gained in defining the Marine Regions Ontology and publishing the Marine Regions gazetteer as LDES (Lonneville 2021), we are iteratively defining the ontologies for representing:

- marine species, whose description is rooted in the definition of taxa and their scientific names, and extends to geographic distribution, scientific literature, ecological traits and other biological information as represented in WoRMS;
- species observations, describing the occurrence of species in space and time, with associated descriptive data as well as biological and environmental measurements, as recorded in EurOBIS.

Our approach to ontology design is driven by the identification of use cases and competency questions, as well as by existing models and data, also involving the community through workshops, webinars and co-creation sessions. Existing domain ontologies and reference data models, such as Biodiversity Information Standards' (TDWG) Darwin Core, the Taxonomic Concept Schema, Bioschema's profiles for taxa and taxon names, the OpenBioDiv ontology, and the Catalogue of Life Data Package (CoLDP), were considered for the identification of ODPs and reuse, particularly through ontology semantic alignments. As fostering consensus on data semantics is a critical step toward broad acceptance and adoption, semantic assets are incrementally published on a dedicated GitHub repository*[5], involving again the community through public review processes.

MAREGRAPH will then enable the production of linked open datasets where the data from WoRMS, Marine Regions and EurOBIS are seamlessly integrated and interlinked in a unified KG that can be further enriched and linked with data from other initiatives where marine biodiversity is the focus.

Keywords

WoRMS, EurOBIS, Linked Open Data, taxa, scientific names, species occurrences, Linked Data Event Streams

Presenting author

Alessandro Russo

Presented at

SPNHC-TDWG 2024

Funding program

Co-funded by the European Union – Digital Europe Programme (DIGITAL)

Author contributions

All authors contributed equally

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Lonneville B, et al. (2021) Publishing the Marine Regions Gazetteer as a Linked Data Event Stream. In: Algergawy A, et al. (Ed.) S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity. URL: https://ceur-ws.org/Vol-2969/paper8-s4biodiv.pdf
- Presutti V, et al. (2009) eXtreme Design with Content Ontology Design Patterns. In: Blomqvist E, et al. (Ed.) Proceedings of the 2009 International Conference on Ontology Patterns. URL: https://ceur-ws.org/Vol-516/pap21.pdf

Endnotes

*1 https://www.marinespecies.org

*2 https://www.marineregions.org

*3 https://www.eurobis.org

*4 which oversees the involvement of the Flanders Marine Institute (VLIZ), the Institute of Cognitive Sciences and Technologies of the Italian National Research Council (CNR-ISTC), Digital Flanders (the digital agency of the Flemish government in Belgium), and imec's Internet Technology and Data Science Lab (IDLab) research group

*5 https://github.com/MareGraph-EU/assets